

Cell Tissue Res (2014) 356:495–505
DOI 10.1007/s00441-014-1855-3

REVIEW

Cis-regulatory variation: significance in biomedicine and evolution

Simon Friedensohn · Ritwick Sawarkar

Received: 8 January 2014 / Accepted: 19 February 2014 / Published online: 18 April 2014
© Springer-Verlag Berlin Heidelberg 2014

Abstract *Cis*-regulatory regions (CRR) control gene expression and chromatin modifications. Genetic variation at CRR in individuals across a population contributes to phenotypic differences of biomedical relevance. This standing variation is important for personalized genomic medicine as well as for adaptive evolution and speciation. This review focuses on genetic variation at CRR, its influence on chromatin, gene expression, and ultimately disease phenotypes. In addition, we summarize our understanding of how this variation may contribute to evolution. Recent technological and computational advances have accelerated research in the direction of personalized medicine, combining strengths of molecular biology and genomics. This will pave new ways to understand how CRR variation affects phenotypes and chart out possible avenues of intervention.

Keywords *cis*-regulatory regions · Chromatin · Genetic variation · GWAS · Evolution

Introduction

How individual genotypes relate to phenotypes is one of the key questions in biology today. Humans within the same population show substantial phenotypic variation in terms of physical properties, behavior, and susceptibility to diseases. This is largely attributed to the genetic variation and environmental differences during growth and development. With the

advent of individual genome sequencing, it should be possible to associate genetic variation with certain phenotypic traits, more importantly those of biomedical relevance. Indeed, personalized medicine promises to customize medical decisions tailored to the genome sequence information of the patient. Thus, the challenge of modern molecular biology and medicine is to make critical predictions about an individual's health based on his/her entire genome sequence. While we are far from making such a causal link, new approaches associate genetic variation within a population to differences in biomedical profiles of individuals.

Genome-wide association (GWA) studies typically compare the probability of certain genetic variants occurring in control and diseased groups to infer if these variants are linked with susceptibility to the disease. Such unbiased studies have been very successful in documenting risk loci associated with diseases like cancers, stroke, and coronary heart conditions, as well as pharmacological traits such as drug efficacy and adverse responses to drugs (Ku et al. 2010; Manolio 2013; McCarthy et al. 2008). While GWA studies only suggest possible loci involved, mechanistic work in disease models is required to firmly establish the causality between genetic variation at these loci and disease susceptibility. About 90 % of loci associated with human diseases fall in non-coding regions of the genome (Hindorff et al. 2009; Maurano et al. 2012; Nicolae et al. 2010). How do the variations in these regions control complex diseases without coding for proteins themselves? Work over the last few decades has elegantly shown that non-coding intergenic regions of the genome are vital in deciding the space, the time, and the quantity of gene expression, and hence protein abundance. The significance of genetic variation at these *cis*-regulatory regions (CRR) is the central theme of this review. We set out by defining the key features of CRR and how they influence the surrounding chromatin architecture and cellular properties. We will then critically evaluate a few instances of medically relevant CRR

S. Friedensohn · R. Sawarkar
Department of Biosystems Science and Engineering, ETH Zurich,
Mattenstrasse 26, 4058 Basel, Switzerland

R. Sawarkar (✉)
Max Planck Institute of Immunobiology and Epigenetics,
79108 Freiburg im Breisgau, Germany
e-mail: sawarkar@ie-freiburg.mpg.de

and their link to chromatin landscape. We discuss how variation at CRR may have played an important role in evolution of new species. Finally, the recent excitement in linking genetic and chromatin variation within a population is summarized, highlighting the tremendous potential of whole-genome sequencing efforts.

CRR, chromatin and transcriptional regulation

Since the discovery of *cis*-regulatory control of transcription by Jacob and Monod half a century ago, we have learned a large number of molecular rules that govern gene expression. While many genomes have extensive intergenic regions, not all of these regions regulate gene activity. How do we identify CRR in the large gene-deserts of the genomes to systematically annotate them? At least four criteria have been used in order to prioritize genomic regions that may have a strong functional role: (1) the presence of clustered TF binding site motifs as inferred by sequence analysis, (2) actual occupancy by chromatin proteins as shown by genome-wide chromatin immunoprecipitation (ChIP) experiments either in cell lines or tissues, (3) biochemical signatures of gene regulation such as histone marks, and (4) regions that are under evolutionary constraint identified by comparing syntenic DNA sequences of related species. Additionally, recent studies have performed unbiased screens for regulatory regions that control gene expression, highlighting enhancers that had been hitherto unknown (Arnold et al. 2013). CRR control various aspects of genomic regulation including transcriptional and post-transcriptional gene expression, RNA splicing, and DNA replication timing. Genetic variation in UTRs or exon/intron junctions will likely affect mRNA stability via miRNA binding and alternative splicing, respectively. Additionally, polymorphisms at DNA replication origins may alter the frequency of DNA replication from those sites. Several studies have reported genetic variation in human population at such sites, even in connection with diseases (Graham et al. 2007; Haas et al. 2012; Rademakers et al. 2008; Wang et al. 2008). This review will focus, however, on the genetic variation at transcriptional control elements like promoters and enhancers.

Transcription factors (TFs) typically bind specific DNA sequences in regulatory regions such as enhancers and facilitate the regulation of target genes by chromatin modifications and influencing RNA polymerase activity. Chromatin forms an important regulatory layer between DNA-binding TFs and gene regulation. More importantly, chromatin integrates DNA-encoded genetic information with heritable environmental influences (epigenetic information). Chromatin at functionally important and active regions such as enhancers and promoters is not highly compacted by nucleosomes and is termed as “open chromatin”. These regions are sensitive to digestion by DNase, and hence are referred to as DNase hypersensitive sites (DHS). Post-translational modifications of promoter-

associated histones such as trimethylation at lysine 27 on histone H3 (H3K27) and H3K4 have been linked with gene repression and activation, respectively. In some cases, both these marks appear simultaneously at promoters representing a poised state of bivalency; these promoters are thought to be ready for activation upon differentiation and are likely to be subjected to the effects of genetic variation (see below). H3K27 and H3K4 trimethylation marks are set by Polycomb and trithorax chromatin-modifying systems (Sawarkar and Paro 2010). On the other hand, several acetyltransferases also act on H3K27 and antagonize the repressive methylation mark set by the Polycomb system. In contrast to promoters, enhancers exhibit a more restricted mark of H3K4 monomethylation and are thought to show chromatin-mediated regulation. Additionally, H3K27 acetylation mark indicates an active enhancer and exhibits dynamic control (Calo and Wysocka 2013). Moreover, CRR exhibit methylation at CpG, typically correlated with gene repression. Often decorated with specific chromatin modifications, CRR variation may even alter the responsiveness to external cues by changing the accessibility of chromatin to environment-regulated TFs. Genetic variation at CRR may influence the surrounding chromatin composition and gene activity at steady state or under environmental dynamics, and ultimately the phenotype at large (Fig. 1). Very recent work has provided empirical evidence on a global scale for this flow of information from variation in DNA sequences to differences in gene activity and surrounding chromatin. Whether chromatin plays an instructive role in this process or is merely consequential to the sequence changes is still a matter of debate.

Standing genetic polymorphism at CRR and consequent chromatin variation

That CRR play a crucial role in directing gene activity may suggest a low genetic variation associated with CRR within a population, just like the coding variation. However, unlike protein-coding variation, CRR variants are likely to affect only a small set of target genes and hence may exhibit only mild deleterious effects. The effect of variation in protein-coding sequences is easy to comprehend in terms of changes in amino acid sequence, and possible consequences on binding, enzymatic activity, or signal transduction. In exceptional cases, variation in the coding regions might also be constrained by the dual use of exonic regions as CRRs (Stergachis et al. 2013). However, variation in CRR is difficult to assess, as the genome-wide annotation of CRR is still coarse-grained with uncertain TF motif definitions and unknown boundaries. To get a full spectrum of the standing genetic variation at CRR including SNPs, insertions, and deletions, at least three aspects have to be considered. First, full genome sequences of sufficient quality from several unrelated individuals within a population need to be examined. Exome sequencing restricted to exons (Bamshad et al. 2011), mainly coding parts of the genome, is now being

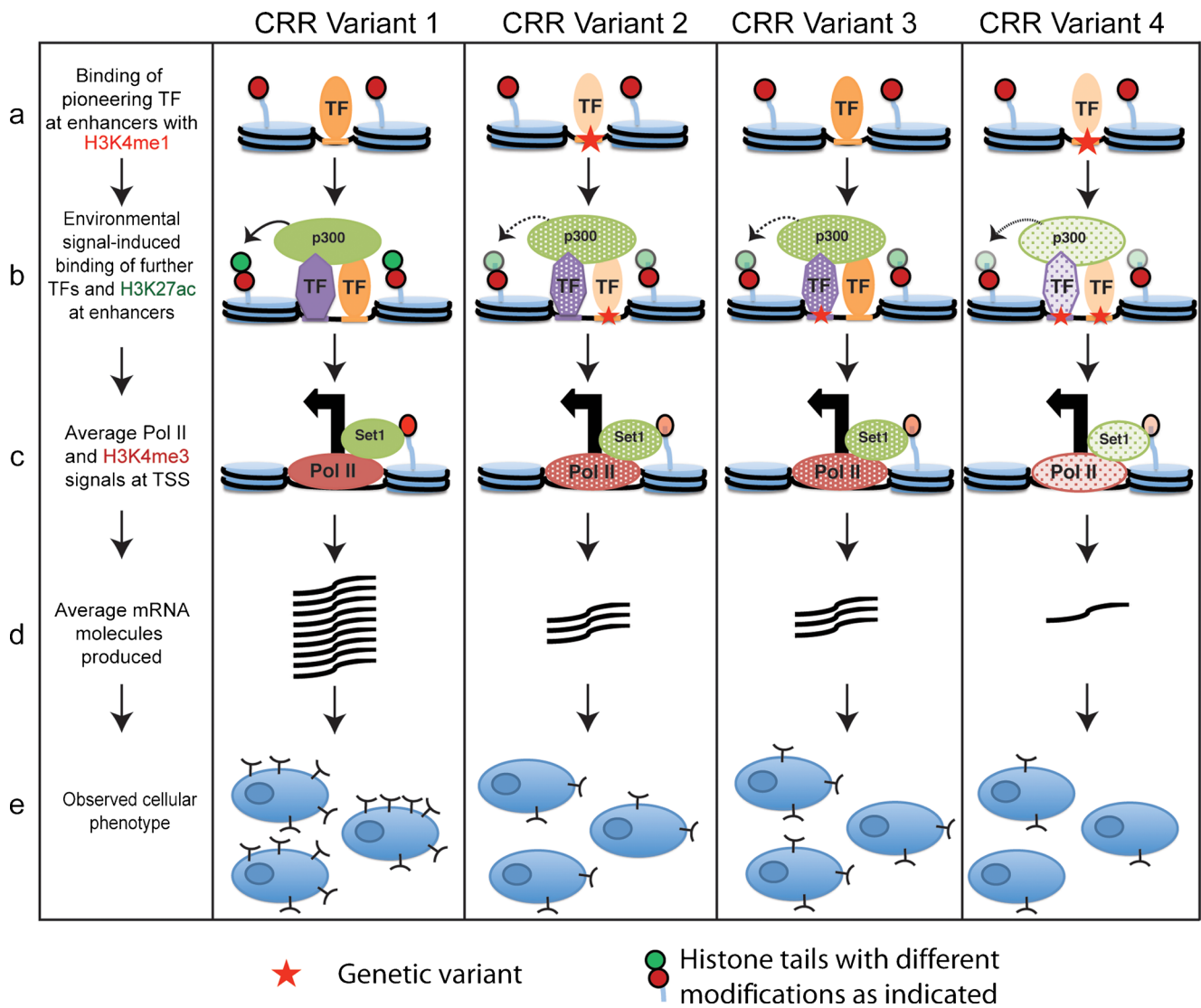


Fig. 1 Linking genotypes with cellular phenotypes via chromatin. *a* Some transcription factors (such as PU.1 or AP1) act as pioneering factors at H3K4me1/2-decorated enhancer sites and induce chromatin modifications. Genetic variants disrupting their cognate DNA-motif (red star) reduce binding of the respective transcription factor, shown here in a lighter shade of orange (e.g., CRR variant 2 and 4). *b* After opening of the chromatin, further transcription factors can bind in a collaborative manner under the influence of extracellular environment. Histone acetyltransferases, like p300, get recruited to the activated enhancer sites. Reduced binding of the initial transcription factor decreases the average ChIP-signal of H3K27ac at the mutated locus, shown here as dotted ovals (variant 2). In a similar

manner, mutations in the motifs belonging to downstream transcription factors also decrease the observed ChIP-signal (variant 3). Combined effects of both mutations are even stronger due to the hierarchical nature of the process (variant 4). *c* Variation in the activity of the upstream enhancer results in variation in Pol II recruitment and trimethylation of H3K4 at transcription start sites (TSS). Again, differential ChIP-signals can be observed for all the cases, consequently leading to dissimilar RNA levels (*d*). *e* The underlying genetic changes lead to variation of a quantitative trait, in this example, the expression of a cell surface receptor. However, the process might be buffered at several post-transcriptional stages, thereby fixing the natural range of the quantitative trait

substituted by whole genome sequencing, allowing CRR variation to surface. An initial plan of fully sequencing 1,000 human genomes from different ethnicities/races is bearing fruits (<http://www.1000genomes.org/>). Second, a detailed understanding of the CRR within the genome is required. Large consortia like ENCODE facilitate systematic annotation of CRR within the human genome by analyzing several chromatin marks and TF binding in different human cell types (Bernstein et al. 2012). Third, knowledge of variation

that can cause quantitative changes in gene expression in defined cell types [called expression quantitative trait loci (eQTLs)] will be useful in discerning causality (Brown et al. 2013; Gilad et al. 2008; Majewski and Pastinen 2011; Morley et al. 2004). Combined whole genome and RNA sequencing of mother–father–child trios is now becoming common allowing an assessment of the inheritance patterns of chromatin variation. Thus, the stage is set to functionally unravel CRR-associated genetic variation within human populations.

The first results of the 1,000-genomes project clearly show that each of the studied individuals carried hundreds of non-coding variants at conserved sites (Abecasis et al. 2012). Many of these variants disrupt TF motifs, and may have functional consequences. Most individual variant alleles were at low frequency in human populations, possibly due to recent population growth and weak selection, just as seen for protein-coding variation (Tennessen et al. 2012). Genome sequences of people from different ethnic backgrounds provide a unique opportunity for studying natural selection and fixation of alleles in human populations. While CRR variation may affect protein amounts within a cell, variation in protein sequences, especially that of TFs, may affect their binding to DNA. Not surprisingly then, studies have found that CRR and coding variation can influence each other such that an effect of a coding variation is contingent upon a CRR variant (Lappalainen et al. 2011).

The large genetic variation at CRR found in human populations hints at possible chromatin variation in *cis*, i.e. changes in chromatin properties in an allele-specific manner. Such an exciting analysis is technically challenging owing to the amounts of samples required to assess chromatin properties genome-wide. In the last few years this has become possible due to the establishment of lymphoblastoid cell lines from people of distinct geographical and racial backgrounds (Frazer et al. 2007; HapMap-Consortium 2003). Additionally, some cell lines came from parents–offspring trios, facilitating analysis of inheritance. These cell lines are an unparalleled resource for which genotype, gene expression, chromatin landscape, and pharmacogenomic information has been made available (Dixon et al. 2007; Jones et al. 2007; Morley et al. 2004; Spielman et al. 2007; Storey et al. 2007; Stranger et al. 2007; Zhang et al. 2008). Moreover, the cell-type-specific effect of CRR variation is also being analyzed within the GenCord project (Dimas et al. 2009). Using these resources, several recent studies set out to decipher fundamental principles of how genetic variation can affect gene expression and chromatin modifications. Integrative genome-wide analyses of DNA sequence variation combined with differences in gene expression and several chromatin properties in these cell lines were performed. A variety of chromatin-related factors were tested, namely DNase I hypersensitivity that indicates open chromatin, several histone modifications marking active and repressed chromatin, RNA pol II activity, occupancy of TF, and selected chromatin proteins. In addition to gene expression variation, all the studies confirmed the existence of large chromatin differences across individuals. This included variation in DNase hypersensitivity, DNA methylation, and active and repressive chromatin marks at promoters and enhancers (Bell et al. 2011; Degner et al. 2012; Gutierrez-Arcelus et al. 2013; Kasowski et al. 2013; Kilpinen et al. 2013; McVicker et al. 2013). Such chromatin variation was stable over generations of cultured cells, and was strongly associated with the

nearby genetic variation at CRR. Enhancers and promoters that exhibit both active and repressive chromatin marks (bivalent regions) show large variation in different histone marks when compared with their active or inactive counterparts (Kasowski et al. 2013). Moreover, all the studies converged on motif-disrupting TF binding site variation that correlated with changes in nearby chromatin marks, suggesting that TF binding is instructive in setting chromatin marks. For example, an NF- κ B motif in the genome showed genetic variation that strongly correlated with differential NF- κ B binding and DNase I hypersensitivity (Degner et al. 2012). Moreover, histone marks at multiple places could be associated with a single DNA variant (Fig. 1; McVicker et al. 2013), possibly via chromosomal contacts like enhancer–promoter looping. Conversely, variation at multiple TFs can correlate with a single histone mark: PU.1 and Myc binding correlates with H3K27Ac (Kilpinen et al. 2013). Interestingly, studies have shown that different TF binding events can cooperate with each other such that variation in one motif affects binding of a distinct TF nearby (Karczewski et al. 2011; Kilpinen et al. 2013), thus affecting chromatin.

Two issues have to be kept in mind while interpreting these data. First, the individual cell lines, though representative of the original genomes, show different growth rates in culture due to unknown reasons (Stark et al. 2010). This may affect gene expression and chromatin. Second, the association of chromatin variation with TF motif changes may not reflect a direct biochemical role of TFs in modifying local chromatin. The differences in chromatin may well be a consequence of gene expression changes set up by CRR variation. Nonetheless, these studies provide the much-awaited empirical data showing that the genetic variation within human populations affects gene expression and chromatin in ways that can be predicted. Most importantly, the studies put TFs at the centre-stage of population genetics linking genetic variation with gene regulation and chromatin. The concept of TF binding to variant enhancers leading to differential gene expression has recently been confirmed using natural genetic variation in mouse strains (Heinz et al. 2013). This underscores the huge impact enhancer variation can have on chromatin and ultimately phenotype. The picture is even more complex if one considers several tissue types with distinct enhancers.

CRR variation and buffering

Although TF motif variation affects nearby chromatin profiles and even nascent transcription, steady-state mRNA levels have been shown to exhibit much less variation (Kasowski et al. 2013; Kilpinen et al. 2013). This could either mean that there is non-consequential genetic variation or that there are ways in which variation can be buffered or compensated for. In some cases, multiple TF motifs occurring within CRR may

allow motif-disrupting or motif-generating variation at a few local sites without altering the regulatory capacity. Additionally, a gene may be controlled by several enhancers and hence simultaneous alterations at multiple enhancers could be required for changes in gene expression. While variation at a single enhancer may cause local chromatin changes, it may not affect the target gene expression unless other enhancers also show variation. In other cases, this compensatory mechanism can be post-transcriptional since the large variability in nascent transcription across human individuals is not reflected in the variation of mRNA abundance. The simplest mechanism is temporal averaging of mRNA produced by a locus owing to the longer half-life of mRNA molecules compared with the transcriptional event per se. An alternative hypothesis proposes that special buffering mechanisms operate at chromatin (Spivakov et al. 2012). Molecular chaperones such as heat-shock protein 90 (Hsp90) have been thought of as a general buffer against protein-coding variation as they can bind to and fold minor variants of proteins (Jarosz et al. 2010). Given Hsp90 is also located at chromatin (Sawarkar and Paro 2013; Sawarkar et al. 2012), it may possibly buffer CRR variation by allowing TFs to bind to the variant DNA with an almost equal affinity. An exciting outcome of this is the possibility that the variation can become responsive to environmental stress: Hsp90's buffering ability may be reduced during conditions of stress. Consequently, CRR variation may result in differential TF binding only upon stress, leading to chromatin and expression differences. Theoretically, such a mechanism also explains gene–environment interactions that are vital to population genetic studies.

CRR variation and human health

Studies from model organisms like *Drosophila* have shown that single mutations in CRR can cause severe developmental abnormalities (e.g., Gyurkovics et al. 1990). However, this does not directly imply that population-wide genetic variation at CRR may be causally associated with human diseases. As compared to coding regions, several reasons have prevented CRR from being empirically correlated with diseases in human populations. Practically, the majority of genome-wide studies focus on exomes that are mostly representative of the protein-coding fraction of the genome (Bamshad et al. 2011), excluding CRR from consideration. Additionally, global GWA studies often focus on functionally validating the effects of protein-coding variants since they are rather straightforward. Polymorphisms within CRR are challenging to interpret, as it is difficult to infer whether a particular variant will affect any function. Finally, CRR variation affecting enhancers is likely to have a highly tissue-specific defect. Unless the right disease model or cell-type is employed for validation, functional consequences of CRR variation can be easily missed. Nonetheless, unbiased GWA studies have

indicated that 80–90 % of the disease QTLs fall outside coding genes (Hindorff et al. 2009; Maurano et al. 2012; Nicolae et al. 2010). Interestingly, many of these loci coincide with DNase-sensitive regions, indicating that active chromatin harbors a large fraction of disease-causing polymorphic sites (Maurano et al. 2012). Several new studies have taken the bold step of molecularly linking these CRR variants with chromatin and gene expression alterations as summarized in the following section and in Table 1.

Similar to protein-coding genes, CRR variation can be linked to diseases with classic Mendelian traits such as hemophilia as well as with more complex multifactorial syndromes like diabetes, cancers, and blood cholesterol levels (Dodd et al. 2013; Gaulton et al. 2010; Musunuru et al. 2010; Pomerantz et al. 2009; Stitzel et al. 2010; Weedon et al. 2014). The key issue is to establish a causal association between CRR variants and disease outcomes, as GWA studies only indicate co-occurrences. For a complex syndrome, a molecular phenotype that represents the disease helps settling the case of causality, as shown for hemophilia B. Mutations in the clotting factor IX (F9) lead to monogenic disorder hemophilia B. Several of these causal variants are found in CRR of the F9 gene that disrupt binding sites of C/EBP- α or HNF4 TFs (Funnell and Crossley 2014). These genetic variants most likely alter chromatin and mRNA expression of the clotting factor gene, leading to hemophilia of a certain subtype. In such a case of a monogenic simple trait, discordance between such mutations and disease symptoms might shed light on putative modifier loci. These may include variants within the CRR itself or genes encoding the cognate TFs. GWA studies of fetal hemoglobin levels have pointed to several CRR variants, some which fall near the gene encoding the repressor of fetal haemoglobin, BCL11A. Recent studies have found one such variant that abolishes GATA1 and Tal1 binding site, affecting chromatin and transcription of the repressor BCL11A altering fetal hemoglobin levels (Bauer et al. 2013).

The genetic and environmental contributions to cancer susceptibility are being intensely investigated. GWA studies have allowed empirical realization of this issue and have mapped several genetic variants. In one case, a breast cancer risk allele was shown to harbor a single SNP within a FOXA1 binding motif near a tumor suppressor gene (Cowper-Salazar et al. 2012). The genetic variant altered FOXA1 binding, and consequently changed the cognate gene expression, increasing the susceptibility of carrier individuals to breast cancer. Another study has underlined the importance of p53 binding sites in CRR of KITLG, a gene implicated in various cancers (Zeron-Medina et al. 2013). The variant causes misregulation of KITLG, presumably accompanied by altered chromatin properties, increasing susceptibility to cancers. The mechanism of enhanced risk of cancer associated with this particular genetic variant is not clear, but could involve other components of

Table 1 Recent findings coupling *cis*-regulatory mutations and human diseases

Associated genetic variant	Impact on molecular phenotype and health
Creation of a new C/EBP binding site at the 1p13 locus	Binding of C/EBP transcription factors to a cognate binding site, which was newly created by a CRR variant, modulates expression of <i>SORT1</i> in hepatocytes. This has been shown to directly influence low-density lipoprotein cholesterol (LDL-C) levels, a known risk factor for myocardial infarction (Musunuru et al. 2010).
Disruption of a binding site for STAT1 at the 9p21 locus	Two CRR variants associated with coronary artery disease (CAD) disrupt a predicted STAT1 binding motif. This enhancer changes its associations dependent on interferon- γ signaling. Thus, the risk alleles may become contingent upon inflammatory responses (Harismendy et al. 2011).
Alteration of a FOXA1 binding sites at the 16q12 locus	Several CRR variants that are associated with breast cancer modulate the affinity of the forkhead protein FOXA1 to chromatin. A specific CRR variant could be shown to repress the tumor suppressor <i>TOX3</i> gene via FOXA1 binding (Cowper-Sal.lari et al. 2012).
Disruption of an ETS-binding motif/ E-box element within the intronic region of the <i>GATA2</i> gene	MonoMAC syndrome characterized by mono- and lymphocyte deficiencies is typically associated with reduced function of GATA2. Mutations in the enhancer of the gene, located in an intron, cause GATA2 expression defects and consequently the syndrome (Hsu et al. 2013).
Disruption of an enhancer motif driving <i>TBX5</i> expression during heart development	Patients with Holt–Oram syndrome suffer from impaired heart and hand development before birth. A mutation in an enhancer bound by an unidentified transcription factor affects the expression of <i>TBX5</i> gene during heart development. People homozygous for the motif-disrupting allele show congenital heart defects (Smemo et al. 2012).
Polymorphisms in the near vicinity of a HIF binding motif at the 11q.13.3 locus	Common variants near the binding-motif of the hypoxia-inducible factor (HIF) modulate the activity of the enhancer as well as the allelic expression of <i>CCND1</i> , a known oncogene. GWA studies associated the implicated SNPs with kidney cancer (renal cell carcinoma, RCC) and further results showed the enhancer to be active in an RCC-specific background (Schoedel et al. 2012).
Deletion of an EGR2 enhancer	The transcription factor EGR2 regulates the development of the myelin sheath in the peripheral nervous system. Homozygous deletion of an enhancer, the myelinating Schwann cell element (MSE), controlling the expression of <i>EGR2</i> results in the lack of myelin and subsequently causes congenital neuropathy (Funalot et al. 2012).
Six recessive mutations that abolish the activity of an enhancer regulating the <i>PTF1A</i> gene.	Patients suffering from non-syndromic pancreatic agenesis fail to undergo normal development of the pancreas. This leads to neonatal diabetes mellitus and exocrine pancreatic insufficiency. The <i>PTF1A</i> gene encodes a TF that regulates pancreatic development and is itself controlled by a conserved enhancer region 25 kb downstream of its promoter. Disruption or deletion of several binding motifs within this conserved region prevents binding of FOXA2, PDX1, and an unknown TF resulting in an inactivation of the <i>PTF1A</i> gene (Weedon et al. 2014).

chromatin/transcriptional regulation. Similarly, the molecular underpinnings of asthma susceptibility have implicated a combination of genetic variation and DNA methylation at the Interleukin 4 receptor gene (Soto-Ramirez et al. 2013). Such studies, when performed genome-wide, will likely lead to an improved understanding of the gene–environment interaction involving epigenetic marks. Additional studies are summarized in Table 1.

Most studies prioritize non-coding variant candidates in GWA studies by comparing conservation of surrounding sequences and known ChIP/histone marks. Consequently, these studies look for loss-of-function variation, e.g., when a highly conserved TF motif is mutated. However, a variant could also create a new TF binding motif in a region that shows little evolutionary conservation or histone marks. A variant that creates GATA-1 TF motif has been reported near the alpha globin gene (De Gobbi et al. 2006). This promoter-like

element, in a region that is otherwise neutral in its regulatory potential, recruits several erythroid lineage TFs. The resultant chromatin structure appears to restrict activation of the cognate alpha globin gene leading to thalassemia. This highlights the importance of taking an unbiased approach in interpreting variants from GWA studies, especially when looking at CRR.

CRR variation and evolution

Molecular evolution is typically studied across closely related species, which gives insight into how genomic sequences change over evolutionary time scales. It can also identify loci that may have actually contributed to the process of divergence. How important are the changes in CRR to adaptive evolution? Work in a variety of species has indicated that CRR have the potential to aid species divergence. However, their relative contribution to molecular evolution compared with coding

sequences is still debated (Hoekstra and Coyne 2007; Linnen et al. 2013). CRR changes tend to be less pleiotropic with milder fitness effects as compared to coding sequences, since the former controls only a small set of genes directly, typically one or two. Also, variation at tissue-specific enhancers is likely to affect the phenotype of that specific tissue, unlike protein variants that affect all the tissues in which that protein is expressed. Therefore, it has been argued that natural selection operates differently (and possibly more effectively) on CRR-based genetic variation resulting in 'fine-tuning' of quantitative traits (Wray 2007).

Changes in the regulatory landscape have been linked to the evolution of vertebrates. A recent study found three 'waves' of regulatory gains during the origin of vertebrates. The first gains occurred in regions regulating key transcription factors and developmental genes, the second wave included gains at genes linked to extracellular signaling, and the last changes happened in CRR of genes encoding post-translational protein modifiers (Lowe et al. 2011). In a similar manner, adaptive, species-specific gains of DHS sites have been connected with the origin of humans. Changes in the chromatin accessibility of genes among primates are correlated to differential gene expression patterns and ultimately to different phenotypic traits (Shibata et al. 2012). Elegant studies have linked development of wing spots in closely related *Drosophila* species to causal variants at tissue-specific enhancers of certain genes (Arnoult et al. 2013). Yet another example for the importance of CRR in evolution is the development of new CTCF binding sites in several *Drosophila* species (Ni et al. 2012). CTCF binding sites are well-known insulator elements and, although the CTCF binding motif is highly preserved, the binding behavior of CTCF differs among these *Drosophila* species. The birth of new CTCF binding sites is subject to purifying selection and correlates with species-specific transcription profiles in adjacent genes and the birth of new genes. This raises the question whether changes in the chromatin architecture are needed for the development of new traits, and hence new species.

While genomes of several related species have been sequenced, different individuals within the same species are only now being intensely investigated: e.g., the 1,000-genome project documenting the variation among humans. Within-species standing variation is the real substrate upon which natural selection operates. Thus, by comparing individual genomes inhabiting different geographic niches, it is possible to connect selection forces with genetic determinants. An example of adaptive selection at a TF binding site is the human DARC locus where a causal SNP in a CRR disrupts binding of GATA1, leading to altered expression of the gene specifically in erythrocytes. Since DARC is a receptor for the entry of malarial parasite, differential expression of this protein has an impact on resistance against malaria. The binding sites of GATA1 have been shown to exhibit a strong amount of positive

selection in human evolution corresponding to the prevalence of malaria within ecological niches (Arbiza et al. 2013). What might then be a connection between conservation of CRR across different species and the genetic variation at these CRRs across individuals of the same species? Loss of TF binding sites in *Drosophila* species may result from the gene expression variation within species. For example, geographically isolated strains may undergo drift or adaptive selection on gene expression. These changes may then be fixed as a loss of TF motifs during speciation. Both these phenomena have been observed in *Drosophila* (Hutter et al. 2008; Kim et al. 2009), but their empirical relationship with each other remains unclear. Thus, genetic variation at CRR may already exist within a population that can be fixed during speciation, either by adaptive selection forces or by chance.

Theoretically, genetic variation at CRR may impact the intensity of binding of TFs if the variation modifies the cognate core motif. However, studies have indicated that variation in the core motif does not always modify the TF binding intensity. Rather, binding behavior depends on the sequence context in which the motif exists. For example, polymorphism at conserved CTCF binding sites has much less effect on CTCF ChIP signals than similar mutational changes in binding sites that were not conserved among species (Spivakov et al. 2012). Furthermore, these effects could not be explained by additional CTCF motifs in the near vicinity, as 96 % of the surveyed loci only contained a single binding site. These findings suggest additional buffering mechanisms, e.g., some possibly mediated by Hsp90 (see above), that stabilize the binding of transcription factors at conserved regions. An exciting possibility is that functional genetic variation at heat shock loci might impact speciation in times of stress. Indeed, changes in core promoter sequences of these genes are linked with varying expression of heat shock proteins in different species of the fly *Liriomyza* (Chen et al. 2011).

The role of chromatin in linking CRR variation to phenotype

As summarized above, many studies have demonstrated differences in chromatin properties owing to changes in CRR sequences that correlate with diseases or are linked with speciation. The critical question is whether chromatin plays an instructive role in such instances. This issue touches on an important debate in epigenetics as to the relevance of heritable chromatin changes in establishing phenotypic variation (Ptashne 2013; Ptashne et al. 2010). In most of the examples cited throughout this text, TF binding site variation, and hence DNA sequence differences, is the cause of chromatin and gene expression changes. These studies do not even address the importance of chromatin, as they investigate genetic changes as a basis for phenotypic variation. Recent studies have employed clever experiments to dissect this conundrum. It is known that the heat-shock transcription factor (HSF), like

most other TFs, only binds to a subset of binding motifs in the genome. One of the explanations for this observation is that motifs in certain genomic regions are not available for TF binding due to the local chromatin architecture. Artificially increasing DNA accessibility in these regions allows TF binding (Guertin and Lis 2010). These observations suggest that chromatin may play a crucial role in deciding whether or not a TF can bind to its cognate motif. How is local chromatin architecture controlled? Possibly this is via heritable epigenetic systems that can be activated by developmental or environmental signals (Cavalli and Paro 1998). Interestingly, work in plants has shown that variation in DNA methylation linked with certain traits can be transmitted across generations, even in the absence of DNA sequence variation (Johannes et al. 2009). This underscores the plausibility of heritable epigenetic variation playing an instructive role in gene regulation and phenotypes. An even more direct evidence that chromatin modifications can play an instructive role in gene expression comes from recent epigenome editing studies. By experimentally manipulating local chromatin over CRRs, cognate gene expression could be modified (Koneremann et al. 2013; Maeder et al. 2013; Mendenhall et al. 2013). Such studies can provide a direct evidence of the role of chromatin marks in influencing transcription: DNA and histone modifications need not be a passive consequence of RNA pol II activity in all these cases. However, these studies are carried out in uniform genetic backgrounds, thus cannot assess the relevance of chromatin in the context of natural CRR variation. To gain more insights into such roles of chromatin, three model systems with controlled genetic and/or environmental backgrounds are being used. First, research on monozygotic twins allows us to understand how different CRR variants are subjected to environmental influences in identical pairs (Bell and Spector 2011). Second, the follow-up studies of the famous case of Dutch famine provides an opportunity to evaluate how prenatal environment can have lasting influence in a real-life situation, presumably via nongenetic alterations (Lumey et al. 2007). Finally, population genetics of model organisms like *D. melanogaster* is a rich source of information on chromatin, as both genetic and environmental backgrounds can be controlled (Mackay et al. 2012). Harnessing genetic variation in natural populations will pave a way for critically establishing contributions of chromatin and genetic factors. These studies are beginning to unravel the complex interplay between chromatin and CRR variation.

Concluding remarks

It is becoming increasingly clear that CRR show substantial genetic variation within members of a population. Much of the variation affects expression of genes via its impact on TF

binding and consequent effects on chromatin, thus affecting human health and susceptibility to diseases. Considerable evidence suggests the involvement of CRR variation in natural selection and adaptation. Our molecular comprehension of gene regulation is paving new ways for understanding how CRR variation can impact biomedicine, ecology, and evolutionary biology. While most studies show correlations of CRR variation and phenotypic differences, a key issue is establishing a causal association and hence predicting the outcome of CRR variation (Cooper and Shendure 2011). Aspects such as the complexity of gene regulation, redundancy and interaction between *cis* and *trans* elements have to be taken into account. Gene–environment interactions, possibly via heritable epigenetic modifications, should also be considered in future studies. Incomplete penetrance of most CRR variants, lack of our knowledge on modifier loci, and the possibility of buffering mechanisms complicate the analyses. However, modern systems biology and computational tools (Evangelou and Ioannidis 2013; Patnala et al. 2013), large sets of population genomic datasets, and ever-decreasing sequencing costs are making genetic variation studies more realistic. In connecting molecular biology with such disparate areas as medicine and evolution, the exciting research on genetic variation truly integrates all of biology.

Acknowledgments The authors thank Renato Paro for scientific discussions; and Jeremy Swann, Eirini Troupouki and an anonymous referee for comments on an earlier draft of this manuscript. R.S. acknowledges support from the Swiss National Foundation (Ambizione grant PZ00P3_142639). The authors apologize to scientists whose work could not be cited due to space limitations.

References

- Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, Handsaker RE, Kang HM, Marth GT, McVean GA (2012) An integrated map of genetic variation from 1,092 human genomes. *Nature* 491:56–65
- Arbiza L, Gronau I, Aksoy BA, Hubisz MJ, Gulko B, Keinan A, Siepel A (2013) Genome-wide inference of natural selection on human transcription factor binding sites. *Nat Genet* 45:723–729
- Arnold CD, Gerlach D, Stelzer C, Boryn LM, Rath M, Stark A (2013) Genome-wide quantitative enhancer activity maps identified by STARR-seq. *Science* 339:1074–1077
- Arnoult L, Su KF, Manoel D, Minervino C, Magrina J, Gompel N, Prud'homme B (2013) Emergence and diversification of fly pigmentation through evolution of a gene regulatory module. *Science* 339:1423–1426
- Bamshad MJ, Ng SB, Bigham AW, Tabor HK, Emond MJ, Nickerson DA, Shendure J (2011) Exome sequencing as a tool for Mendelian disease gene discovery. *Nat Rev Genet* 12:745–755
- Bauer DE, Kamran SC, Lessard S, Xu J, Fujiwara Y, Lin C, Shao Z, Canver MC, Smith EC, Pinello L et al (2013) An Erythroid Enhancer of BCL11A Subject to Genetic Variation Determines Fetal Hemoglobin Level. *Science* 342:253–257
- Bell JT, Spector TD (2011) A twin approach to unraveling epigenetics. *Trends Genet* 27:116–125

- Bell JT, Pai AA, Pickrell JK, Gaffney DJ, Pique-Regi R, Degner JF, Gilad Y, Pritchard JK (2011) DNA methylation patterns associate with genetic and gene expression variation in HapMap cell lines. *Genome Biol* 12:R10
- Bernstein BE, Bimney E, Dunham I, Green ED, Gunter C, Snyder M (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature* 489:57–74
- Brown CD, Mangravite LM, Engelhardt BE (2013) Integrative modeling of eQTLs and *cis*-regulatory elements suggests mechanisms underlying cell type specificity of eQTLs. *PLoS Genet* 9:e1003649
- Calo E, Wysocka J (2013) Modification of enhancer chromatin: what, how, and why? *Mol Cell* 49:825–837
- Cavalli G, Paro R (1998) The *Drosophila* Fab-7 chromosomal element conveys epigenetic inheritance during mitosis and meiosis. *Cell* 93:505–518
- Chen B, Jia T, Ma R, Zhang B, Kang L (2011) Evolution of hsp70 gene expression: a role for changes in AT-richness within promoters. *PLoS ONE* 6:e20308
- Cooper GM, Shendure J (2011) Needles in stacks of needles: finding disease-causal variants in a wealth of genomic data. *Nat Rev Genet* 12:628–640
- Cowper-Salari R, Zhang X, Wright JB, Bailey SD, Cole MD, Eeckhoutte J, Moore JH, Lupien M (2012) Breast cancer risk-associated SNPs modulate the affinity of chromatin for FOXA1 and alter gene expression. *Nat Genet* 44:1191–1198
- De Gobbi M, Viprakasit V, Hughes JR, Fisher C, Buckle VJ, Ayyub H, Gibbons RJ, Vernimmen D, Yoshinaga Y, de Jong P et al (2006) A regulatory SNP causes a human genetic disease by creating a new transcriptional promoter. *Science* 312:1215–1217
- Degner JF, Pai AA, Pique-Regi R, Veyrieras J-B, Gaffney DJ, Pickrell JK, De Leon S, Michelini K, Lewellen N, Crawford GE et al (2012) DNase I sensitivity QTLs are a major determinant of human expression variation. *Nature* 482:390–394
- Dimas AS, Deutsch S, Stranger BE, Montgomery SB, Borel C, Attar-Cohen H, Ingle C, Beazley C, Gutierrez Arcelus M, Sekowska M et al (2009) Common regulatory variation impacts gene expression in a cell type-dependent manner. *Science* 325:1246–1250
- Dixon AL, Liang L, Moffatt MF, Chen W, Heath S, Wong KC, Taylor J, Burnett E, Gut I, Farrell M et al (2007) A genome-wide association study of global gene expression. *Nat Genet* 39:1202–1207
- Dodd AW, Syddall CM, Loughlin J (2013) A rare variant in the osteoarthritis-associated locus GDF5 is functional and reveals a site that can be manipulated to modulate GDF5 expression. *Eur J Hum Genet* 21:517–521
- Evangeliou E, Ioannidis JP (2013) Meta-analysis methods for genome-wide association studies and beyond. *Nat Rev Genet* 14:379–389
- Frazer KA, Ballinger DG, Cox DR, Hinds DA, Stuve LL, Gibbs RA, Belmont JW, Boudreau A, Hardenbol P, Leal SM et al (2007) A second generation human haplotype map of over 3.1 million SNPs. *Nature* 449:851–861
- Funalot B, Topilko P, Arroyo MA, Sefiani A, Hedley-Whyte ET, Yoldi ME, Richard L, Touraille E, Laurichesse M, Khalifa E et al (2012) Homozygous deletion of an EGR2 enhancer in congenital amyelinating neuropathy. *Ann Neurol* 71:719–723
- Funnell APW, Crossley M (2014) Hemophilia B Leyden and once mysterious *cis*-regulatory mutations. *Trends Genet* 30:18–23
- Gaulton KJ, Nanno T, Pasquali L, Simon JM, Giresi PG, Fogarty MP, Panhuis TM, Mieczkowski P, Secchi A, Bosco D et al (2010) A map of open chromatin in human pancreatic islets. *Nat Genet* 42:255–259
- Gilad Y, Rifkin SA, Pritchard JK (2008) Revealing the architecture of gene regulation: the promise of eQTL studies. *Trends Genet* 24:408–415
- Graham RR, Kyogoku C, Sigurdsson S, Vlasova IA, Davies LR, Baechler EC, Plenge RM, Koeuth T, Ortmann WA, Hom G et al (2007) Three functional variants of IFN regulatory factor 5 (IRF5) define risk and protective haplotypes for human lupus. *Proc Natl Acad Sci USA* 104:6758–6763
- Guertin MJ, Lis JT (2010) Chromatin landscape dictates HSF binding to target DNA elements. *PLoS Genet* 6:e1001114
- Gutierrez-Arcelus M, Lappalainen T, Montgomery SB, Buil A, Ongen H, Yurovsky A, Bryois J, Giger T, Romano L, Planchon A et al (2013) Passive and active DNA methylation and the interplay with genetic variation in gene regulation. *Elife* 2:e00523
- Gyurkovics H, Gausz J, Kummer J, Karch F (1990) A new homeotic mutation in the *Drosophila* bithorax complex removes a boundary separating two domains of regulation. *EMBO J* 9:2579–2585
- Haas U, Szakiel G, Laufer SD (2012) MicroRNA-mediated regulation of gene expression is affected by disease-associated SNPs within the 3'-UTR via altered RNA structure. *RNA Biol* 9:924–937
- HapMap-Consortium, T.I. (2003) The International HapMap Project. *Nature* 426:789–796
- Harismendy O, Notani D, Song X, Rahim NG, Tanasa B, Heintzman N, Ren B, Fu XD, Topol EJ, Rosenfeld MG et al (2011) 9p21 DNA variants associated with coronary artery disease impair interferon-gamma signalling response. *Nature* 470:264–268
- Heinz S, Romanoski CE, Benner C, Allison KA, Kaikkonen MU, Orozco LD, Glass CK (2013) Effect of natural genetic variation on enhancer selection and function. *Nature* 503:487–492
- Hindorf LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, Collins FS, Manolio TA (2009) Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci USA* 106:9362–9367
- Hoekstra HE, Coyne JA (2007) The locus of evolution: evo devo and the genetics of adaptation. *Evolution* 61:995–1016
- Hsu AP, Johnson KD, Falcone EL, Sanalkumar R, Sanchez L, Hickstein DD, Cuellar-Rodriguez J, Lemieux JE, Zerbe CS, Bresnick EH et al (2013) GATA2 haploinsufficiency caused by mutations in a conserved intronic element leads to MonoMAC syndrome. *Blood* 121:S3831–S3837
- Hutter S, Saminadin-Peter SS, Stephan W, Parsch J (2008) Gene expression variation in African and European populations of *Drosophila melanogaster*. *Genome Biol* 9:R12
- Jarosz DF, Taipale M, Lindquist S (2010) Protein homeostasis and the phenotypic manifestation of genetic diversity: principles and mechanisms. *Annu Rev Genet* 44:189–216
- Johannes F, Porcher E, Teixeira FK, Saliba-Colombani V, Simon M, Agier N, Bulski A, Albuisson J, Heredia F, Audigier P et al (2009) Assessing the impact of transgenerational epigenetic variation on complex traits. *PLoS Genet* 5:e1000530
- Jones TS, Yang W, Evans WE, Relling MV (2007) Using HapMap tools in pharmacogenomic discovery: the thiopurine methyltransferase polymorphism. *Clin Pharmacol Ther* 81:729–734
- Karczewski KJ, Tatonetti NP, Landt SG, Yang X, Slifer T, Altman RB, Snyder M (2011) Cooperative transcription factor associations discovered using regulatory variation. *Proc Natl Acad Sci USA* 108:13353–13358
- Kasowski M, Kyriazopoulou-Panagiotopoulou S, Grubert F, Zaugg JB, Kundaje A, Liu Y, Boyle AP, Zhang QC, Zakharia F, Spacek DV, et al (2013) Extensive Variation in Chromatin States Across Humans. *Science* 342:750–752
- Kilpinen H, Waszak SM, Gschwind AR, Raghav SK, Witwicki RM, Orioli A, Migliaacca E, Wiederkehr M, Gutierrez-Arcelus M, Panousis N, et al (2013) Coordinated Effects of Sequence Variation on DNA Binding, Chromatin Structure, and Transcription. *Science* 342:744–747
- Kim J, He X, Sinha S (2009) Evolution of regulatory sequences in 12 *Drosophila* species. *PLoS Genet* 5:e1000330
- Konermann S, Brigham MD, Trevino AE, Hsu PD, Heidenreich M, Cong L, Platt RJ, Scott DA, Church GM, Zhang F (2013) Optical control of mammalian endogenous transcription and epigenetic states. *Nature* 500:472–476

- Ku CS, Loy EY, Pawitan Y, Chia KS (2010) The pursuit of genome-wide association studies: where are we now? *J Hum Genet* 55:195–206
- Lappalainen T, Montgomery SB, Nica AC, Dermitzakis ET (2011) Epistatic selection between coding and regulatory variation in human evolution and disease. *Am J Hum Genet* 89:459–463
- Linnen CR, Poh YP, Peterson BK, Barrett RD, Larson JG, Jensen JD, Hoekstra HE (2013) Adaptive evolution of multiple traits through multiple mutations at a single gene. *Science* 339:1312–1316
- Lowe CB, Kellis M, Siepel A, Raney BJ, Clamp M, Salama SR, Kingsley DM, Lindblad-Toh K, Haussler D (2011) Three periods of regulatory innovation during vertebrate evolution. *Science* 333:1019–1024
- Lumey LH, Stein AD, Kahn HS, van der Pal-de Bruin KM, Blauw GJ, Zybert PA, Susser ES (2007) Cohort profile: the Dutch Hunger Winter families study. *Int J Epidemiol* 36:1196–1204
- Mackay TF, Richards S, Stone EA, Barbadilla A, Ayroles JF, Zhu D, Casillas S, Han Y, Magwire MM, Cridland JM et al (2012) The *Drosophila melanogaster* Genetic Reference Panel. *Nature* 482:173–178
- Maeder ML, Angstman JF, Richardson ME, Linder SJ, Cascio VM, Tsai SQ, Ho QH, Sander JD, Reyon D, Bernstein BE et al (2013) Targeted DNA demethylation and activation of endogenous genes using programmable TALE-TET1 fusion proteins. *Nat Biotechnol* 31:1137–1142
- Majewski J, Pastinen T (2011) The study of eQTL variations by RNA-seq: from SNPs to phenotypes. *Trends Genet* 27:72–79
- Manolio TA (2013) Bringing genome-wide association findings into clinical use. *Nat Rev Genet* 14:549–558
- Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, Wang H, Reynolds AP, Sandstrom R, Qu H, Brody J et al (2012) Systematic localization of common disease-associated variation in regulatory DNA. *Science* 337:1190–1195
- McCarthy MI, Abecasis GR, Cardon LR, Goldstein DB, Little J, Ioannidis JP, Hirschhorn JN (2008) Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nat Rev Genet* 9:356–369
- McVicker G, van de Geijn B, Degner JF, Cain CE, Banovich NE, Raj A, Lewellen N, Myrthil M, Gilad Y, Pritchard JK (2013) Identification of Genetic Variants That Affect Histone Modifications in Human Cells. *Science* 342:747–749
- Mendenhall EM, Williamson KE, Reyon D, Zou JY, Ram O, Joung JK, Bernstein BE (2013) Locus-specific editing of histone modifications at endogenous enhancers. *Nat Biotechnol* 31:1133–1136
- Morley M, Molony CM, Weber TM, Devlin JL, Ewens KG, Spielman RS, Cheung VG (2004) Genetic analysis of genome-wide variation in human gene expression. *Nature* 430:743–747
- Musunuru K, Strong A, Frank-Kamenetsky M, Lee NE, Ahfeldt T, Sachs KV, Li X, Li H, Kuperwasser N, Ruda VM et al (2010) From noncoding variant to phenotype via SORT1 at the 1p13 cholesterol locus. *Nature* 466:714–719
- Ni X, Zhang YE, Negre N, Chen S, Long M, White KP (2012) Adaptive evolution and the birth of CTCF binding sites in the *Drosophila* genome. *PLoS Biol* 10:e1001420
- Nicolae DL, Gamazon E, Zhang W, Duan S, Dolan ME, Cox NJ (2010) Trait-associated SNPs are more likely to be eQTLs: annotation to enhance discovery from GWAS. *PLoS Genet* 6:e1000888
- Patnala R, Clements J, Batra J (2013) Candidate gene association studies: a comprehensive guide to useful in silico tools. *BMC Genet* 14:39
- Pomerantz MM, Ahmadiyeh N, Jia L, Herman P, Verzi MP, Doddapaneni H, Beckwith CA, Chan JA, Hills A, Davis M et al (2009) The 8q24 cancer risk variant rs6983267 shows long-range interaction with MYC in colorectal cancer. *Nat Genet* 41:882–884
- Ptashne M (2013) Faddish stuff: epigenetics and the inheritance of acquired characteristics. *FASEB J* 27:1–2
- Ptashne M, Hobert O, Davidson E (2010) Questions over the scientific basis of epigenome project. *Nature* 464:487
- Rademakers R, Eriksen JL, Baker M, Robinson T, Ahmed Z, Lincoln SJ, Finch N, Rutherford NJ, Crook RJ, Josephs KA et al (2008) Common variation in the miR-659 binding-site of GRN is a major risk factor for TDP43-positive frontotemporal dementia. *Hum Mol Genet* 17:3631–3642
- Sawarkar R, Paro R (2010) Interpretation of developmental signaling at chromatin: the Polycomb perspective. *Dev Cell* 19:651–661
- Sawarkar R, Paro R (2013) Hsp90@chromatin.nucleus: an emerging hub of a networker. *Trends Cell Biol* 23:193–201
- Sawarkar R, Sievers C, Paro R (2012) Hsp90 globally targets paused RNA polymerase to regulate gene expression in response to environmental stimuli. *Cell* 149:807–818
- Schoedel J, Bardella C, Sciesielski LK, Brown JM, Pugh CW, Buckle V, Tomlinson IP, Ratcliffe PJ, Mole DR (2012) Common genetic variants at the 11q13.3 renal cancer susceptibility locus influence binding of HIF to an enhancer of cyclin D1 expression. *Nat Genet* 44:420–425
- Shibata Y, Sheffield NC, Fedrigo O, Babbitt CC, Wortham M, Tewari AK, London D, Song L, Lee BK, Iyer VR et al (2012) Extensive evolutionary changes in regulatory element activity during human origins are associated with altered gene expression and positive selection. *PLoS Genet* 8:e1002789
- Smemo S, Campos LC, Moskowitz IP, Krieger JE, Pereira AC, Nobrega MA (2012) Regulatory variation in a TBX5 enhancer leads to isolated congenital heart disease. *Hum Mol Genet* 21:3255–3263
- Soto-Ramirez N, Arshad SH, Holloway JW, Zhang H, Schaubberger E, Ewart S, Patil V, Karmaus W (2013) The interaction of genetic variants and DNA methylation of the interleukin-4 receptor gene increase the risk of asthma at age 18 years. *Clin Epigenet* 5:1
- Spielman RS, Bastone LA, Burdick JT, Morley M, Ewens WJ, Cheung VG (2007) Common genetic variants account for differences in gene expression among ethnic groups. *Nat Genet* 39:226–231
- Spivakov M, Akhtar J, Kheradpour P, Beal K, Girardot C, Koscielny G, Herrero J, Kellis M, Furlong EE, Birney E (2012) Analysis of variation at transcription factor binding sites in *Drosophila* and humans. *Genome Biol* 13:R49
- Stark AL, Zhang W, Zhou T, O'Donnell PH, Beiswanger CM, Huang RS, Cox NJ, Dolan ME (2010) Population differences in the rate of proliferation of international HapMap cell lines. *Am J Hum Genet* 87:829–833
- Stergachis AB, Haugen E, Shafer A, Fu W, Vernot B, Reynolds A, Raubitschek A, Ziegler S, LeProust EM, Akey JM et al (2013) Exonic transcription factor binding directs codon choice and affects protein evolution. *Science* 342:1367–1372
- Stitzel ML, Sethupathy P, Pearson DS, Chines PS, Song L, Erdos MR, Welch R, Parker SC, Boyle AP, Scott LJ et al (2010) Global epigenomic analysis of primary human pancreatic islets provides insights into type 2 diabetes susceptibility loci. *Cell Metab* 12:443–455
- Storey JD, Madeoy J, Strout JL, Wurfel M, Ronald J, Akey JM (2007) Gene-expression variation within and among human populations. *Am J Hum Genet* 80:502–509
- Stranger BE, Nica AC, Forrest MS, Dimas A, Bird CP, Beazley C, Ingle CE, Dunning M, Flicek P, Koller D et al (2007) Population genomics of human gene expression. *Nat Genet* 39:1217–1224
- Tennessen JA, Bigham AW, O'Connor TD, Fu W, Kenny EE, Gravel S, McGee S, Do R, Liu X, Jun G et al (2012) Evolution and functional impact of rare coding variation from deep sequencing of human exomes. *Science* 337:64–69
- Wang G, van der Walt JM, Mayhew G, Li YJ, Zuchner S, Scott WK, Martin ER, Vance JM (2008) Variation in the miRNA-433 binding site of FGF20 confers risk for Parkinson disease by overexpression of alpha-synuclein. *Am J Hum Genet* 82:283–289

- Weedon MN, Cebola I, Patch AM, Flanagan SE, De Franco E, Caswell R, Rodriguez-Segui SA, Shaw-Smith C, Cho CH, Lango Allen H et al (2014) Recessive mutations in a distal PTF1A enhancer cause isolated pancreatic agenesis. *Nat Genet* 46:61–64
- Wray GA (2007) The evolutionary significance of *cis*-regulatory mutations. *Nat Rev Genet* 8:206–216
- Zeron-Medina J, Wang X, Repapi E, Campbell MR, Su D, Castro-Giner F, Davies B, Peterse EFP, Sacilotto N, Walker GJ et al (2013) A Polymorphic p53 Response Element in KIT Ligand Influences Cancer Risk and Has Undergone Natural Selection. *Cell* 155:410–422
- Zhang W, Duan S, Kistner EO, Bleibel WK, Huang RS, Clark TA, Chen TX, Schweitzer AC, Blume JE, Cox NJ et al (2008) Evaluation of genetic variation contributing to differences in gene expression between populations. *Am J Hum Genet* 82: 631–640